Delphes (Offline 2007)

Delphes offered an "intelligent knowledge service." The company's linguistic method focused on making "know how" the foundation of search.

Author's note: This is an unpublished, preliminary draft of a description originally destined for my "Encyclopedia" of Enterprise Search, first published by CMS Watch between 2004 and 2006. Versions of this report have appeared in documents I prepared for financial services clients, and I have referenced the details in this report in my articles for various journal publishers.

In 2007, Delphes Technologies International, the linguistic soul vendor, went out of business. The firm's final marketing angle was to position the Delphes system as a solution to French-speaking organizations information retrieval problems. At its peak, Delphes asserted that it had licensed its technology to more than a dozen companies.

Like other "linguistics-based" content processing systems, Delphes faced long sales cycles and increased marketing costs. Delphes was one of the first enterprise search systems to introduce the concept of "psycho linguistics" into content processing and a "multidisciplinary perspective" for search. Delphes is a case study that shows how an academic approach to information retrieval can culture a massive, complex ecosystem of subsystems. Performance, cost, and maintenance are, it seems, the challenge of vendors who have big ideas but smaller resources.

This information is a rough draft and is frozen.

Introduction

Founded in 1998, Delphes Technologies International commercialized academic research related to indexing content using linguistic technology. The company tapped search expertise from Canadian and US universities. The company's approach drew heavily on academic research. Dr. Anna Marie Di Scuillo, one of the founders of Delphes, is a strong proponent of semantic and linguistic methods applied to information retrieval. The Delphes product nomenclature makes use of the "di" in her name.

Delphes supported English, French, Spanish, and German. Unlike some other late 1990s search systems, Delphes had multi-language capabilities built in.

Delphes, like other "do everything" search systems offered an integrated information system.

Table 1: Delphes DioSMW: A Bird's Eye View

Product Thumbnail	
1 Search Brand	Delphes UNAX and product variants called DioSMW (the enterprise solution), Dio Millennium Series (a less robust version of the system), and DioWeb Series (a hosted solution)
2 OS Supported	Microsoft Windows
3 Est License Fee	Pricing is based on the number of documents the system takes under management. For a collection of one million documents, the estimated fee, including maintenance, is about US$2.2 million
4 Functions	Knowledge extraction, information retrieval, data, information, and knowledge management
5 Claimed Features	Automatic concept expansion, natural language processing via linguistic methods, and universal information access for system users, multi-language support
6 Downsides	The company combines philosophical views of knowledge with enterprise information access
7 Similar To	Autonomy IDOL, Convera RetrievalWare, OpenText LiveLink, Fast Search & Transfer ESP, Verity
Product Close Up	Delphes began as an academic research project. The commercialization of Delphes focused on serving the needs of Francophone organizations in Canada and Europe. The system's computational approach requires significant resources to index backfiles, new content, and changes to already-processed content. the company focuses on giving licensees access to the knowledge within an organization. Search is a key process. The firm's UNAX system provides "digital intelligence" to users. Unlike traditional enterprise search systems, Delphes focuses on a somewhat nebulous concept--knowledge management.

Delphes Technologies offers an intelligent knowledge management system for the enterprise. The system contains search and retrieval, entity identification, key word and concept indexing, analytics, and a host of other knowledge and data management services. Like Autonomy and Fast Search, Delphes offers a licensee a total solution; that is, once installed, licensees can access, analyze, synthesize, and share information with colleagues. Del-

phes edges into Endeca's eCommerce territory with its assertion that the Delphes system can support electronic market places. Anticipating Convera's abrupt shift from enterprise search to niche search, Delphes was one of the first enterprise search vendors to offer support for vertical search. (Vertical search means a constrained collection of content usually related to a topic, a publishing company's content, or a directory.)

Like Verity, Delphes offered consulting and engineering services for its search system as well as for infrastructure operations, financial analysis, and general management consulting.

Management

The senior management of Delphes includes:

- Sylvain Falardeau, Chief Executive Officer and President
- Annie Demers, Vice President of Operations
- Harold Roussel, Chief Technical Officer
- Georges Valade, Vice President of Development
- Dr. Anne Marie Di Sciullo, Vice President of Linguistic Strategy

Clients

The firm's clients include:

- Bell Canada
- Bombardier Transport (Germany)
- Canadian Museum of Civilization Corp.
- Canadian Union of Postal Works
- Centre d'accés a l'information juridique. The network consists of 37 law libraries
- Desjardins Financial Security
- Desjardins Credit Unions Federation
- Fédération des caisses Desjardin
- Industry Canada
- L'Oréal USA
- Québec Bar Association. 20,000 users.
- Québec Department of Finance
- SNC Lavalin. The Delphes system supports 8,000 users

Notice that the company lists one client, Dejardins, multiple times. Is this an attempt to puff up the client list?

The Birth of Digital Intelligence

Delphes embraced a broad vision and the fuzzy jargon of knowledge management. The Delphes "Integrated Information System" white paper explains:[1]

> Knowledge is of the highest value in an organization. Knowledge management is the set of practices and technologies that leverages value-added information, which is now critical for more and more organizations.

Delphes' approach is to manage "knowledge." The idea is that the Delphes system will go beyond connecting employees via an internal network. The key to success, according to Delphes, is that

> it also takes comprehensive and effective information searching, document and knowledge managing.
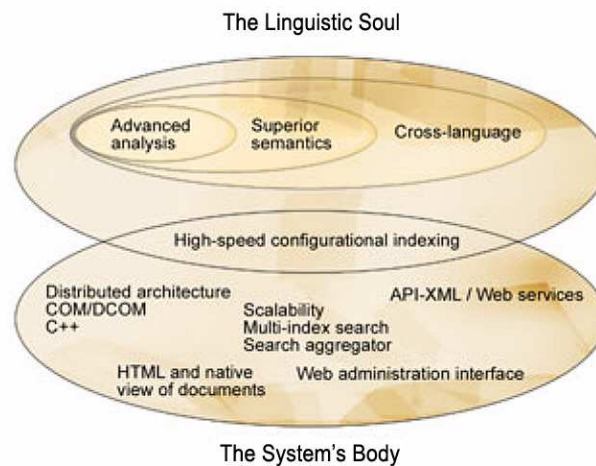
Delphes suggests that software can impart "soul" to search via numerical techniques. Source: Delphes white paper. No date.

The Delphes technology is not connected to a specific benefit like more profit or higher sales. Delphes takes a more philosophical approach to search:

> Knowledge is both tangible and intangible! It is tangible in the sense that it is globally accessible in millions of databases, directories, Web sites. It is intangible in the sense that it is inaccessible if we do not have the necessary tools to quickly and efficiently extract

---

[1.] *The Birth of Digital Intelligence White Paper: Integrated Information System*, published by Delphes, 2003.

4

> the essentials. Here enters Delphes' integrated information system, the ultimate knowledge-sharing system for your organization.

Delphes seeks to impact a "linguistic soul" of its search system. The illustration of "Delphes' Evolving System" is a refreshing departure from block diagrams of search systems.

One important consequence of the Delphes approach is that the company incorporates a "universal grammar." A user can search for a concept and get relevant results, a claim made by other enterprise search systems. A user can enter a query in English and receive results regardless of the language of the documents matching the query.

UNAX

The core of the Delphes system is a Universal Axiomatic Engine dubbed UNAX.[2] Four important operations have been identified.

First, UNAX identifies "abstract structured entities." Delphes calls these configurations. A configuration may be structured sets of characters, morphemes, or words, phrases, or texts. The idea is that the system can imitate the methods used by a person to find information. According to Delphes:

> These activities are brought about in terms of abstract categories, a large subclass of which are not visible by the human perceptual system, but nevertheless are part of the cognitive algebra that makes humans able to manipulate information. UNAX mimics a fundamental feature of the human cognitive system: the ability to process information supported by natural language in terms of the manipulation of abstract configurations and categories.

Second, UNAX preserves relationships between the query and "equivalent expressions.

Third, UNAX includes natural language processing modules. These identify concepts associated with the user's query. Like other linguistic-based systems, UNAX identifies bound phrases such as "white house" and "stock market". The combination of concept expansion and stemming makes UNAX "a morphological powerhouse." Delphes asserts:

> UNAX derives conceptual expansions from the relations between syntactic constituents. The identification of the conceptual relations supported by nominal expressions is central in the system, as the referent (object of a search) is supported mainly by nominal expres-

---

[2] The approach is described in the 2001 paper Di Sciullo, Anna Maria and Fong, Sandiway (2001). "Efficient Parsing for Word Structure". In the Proceedings of the Sixth Natural Language Processing Pacific Rim Symposium. November 27-30, 2001.

sions in natural languages. The ability to associate nominal configurations to conceptual configurations reduces the set of possible interpretations for a given expression on the one hand, and, on the other hand, renders the retrieval and extraction processes more precise.

Finally, UNAX delivers "evolved text search." The idea is that the Delphes system takes a fundamentally different approach to systems which use key word search. Instead of string matching, UNAX delivers a conceptual dimension to information retrieval.

Search and Retrieval

Delphes says that its integrated information system can determine the contextual purpose of words. The indexing method uses UNAX's proprietary configuration method. Delphes analyzes the text, parsing words, phrases, and larger grammatical units. The content is analyzed. Specialized dictionaries are used to identify bound phrases and proper nouns. Dictionaries, however, have to be maintained or their value diminishes over time. Delphes asserts that its methods make it possible to identify the context for indexed elements from a document.

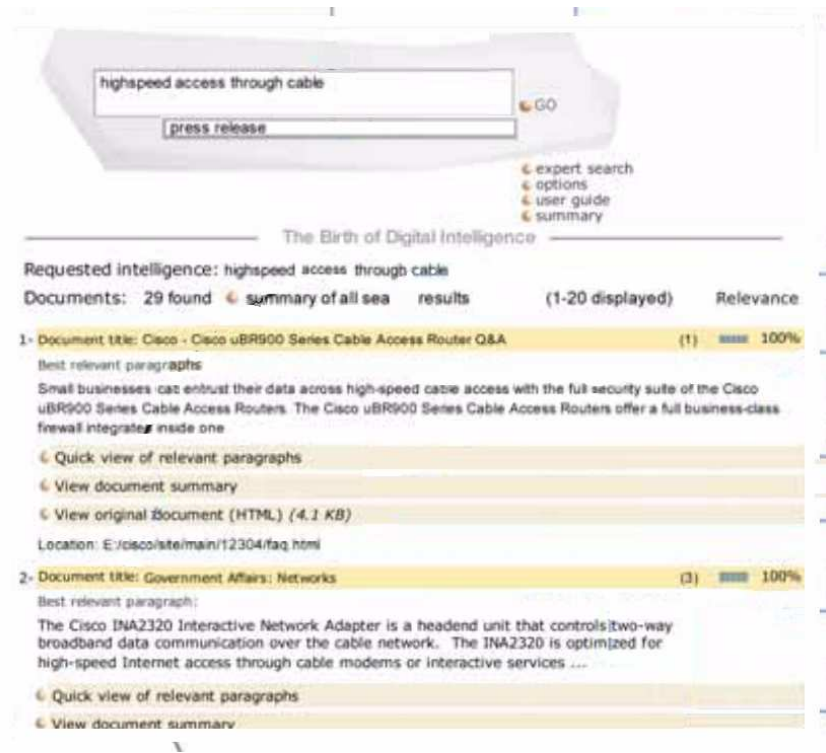## Indexing and Search

Indexing

Both the data and metadata are indexed to allow for efficient retrieval of the extracted information. Indexing can occur on a regular schedule and be limited by document size, date, type, language, section, and URL. These capabilities are enabled through the use of the Universal Axiomatic Engine (UNAX).

Delphes processes the body of the text, also included are annotations, metatags, notes, bookmarks, and titles. Parameters allow for defining the depth of indexing levels, the document's size, date, and type, as well as the language used for indexing. These can be adjusted on a per-user basis.

Delphes allows for indexing multiple batches of content simultaneously. The system can index in parallel multiple sources of information, coming from Web sites as well as intranet servers. Setting up the data acquisition can be performed remotely from any computer with permission to access the administration module.Customized categorization of documents is supported.

The system is capable of processing over 200 file formats, including Lotus Notes. A fee is charged for the Lotus Notes connector, however. In addition to indexing the body of the text, also included are annotations, metatags,

notes, bookmarks, and titles. Parameters allow for defining the depth of indexing levels, the document's size, date, and type, as well as the language used for indexing. These can be adjusted on a per-user basis.

The Delphes interface offers a number of controls. Hot links make it easy to jump from a summary of the results to snippets of text from the source document.

Delphes allows for the indexing and searching of information within PDF files which have been encrypted to prevent unauthorized printing, modification, or reproduction. No additional configuration is required for this module, and upon activation is transparent to the user.

Search

Results are presented dynamically within a web-based interface. Characteristics include:

- Recognizes relevant linguistic concepts within the query;

- Link positioning the user at the right place within the document;

- Summarizes the most pertinent paragraph of each relevant document;

- High-performance relevancy module

- Link accessing the original document

One useful feature is that a hit in a results list takes the user to the specific passage in a document where the relevant information appears.

Delphes asserts that its system performs semantic search. The system can distinguish headings, names, subjects, verbs and complements in order to extract the query's meaning and relate the query to associated concepts

Statistics

Delphes offers a number of different statistical views, presented to an auditor via a web interface:

- Compilation of search queries

- Analysis of search sessions by user

- Results classification by language, category, and IP address

- Reports on automated indexing tasks. These can be exported in a comma separated values (CSV) format for reuse with other analysis tools.

Licensees can use these data to tune relevance via the administrative interface and obtain information about user behaviors.
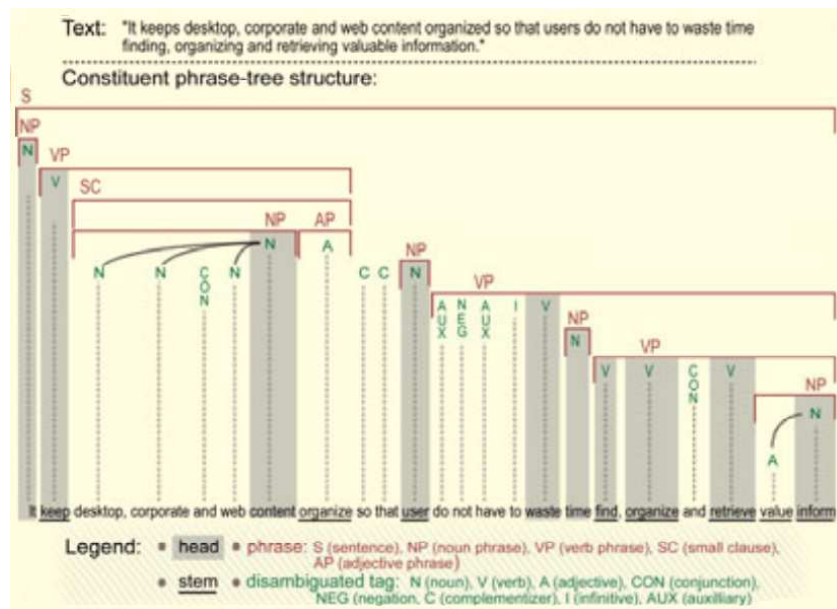
Advanced Functions

Linguistics

Delphes implements a large number of linguistic functions. Linguistic, semantic, and natural language processing require sufficient computational resources.

- Advanced analytical capabilities distinguish a query's related concepts

- Identification of spelling variants. Delphes uses the term *morphological concepts* to describe this operation.

- Spell-checks queries and automatically processes accents. The system can display alternate spellings automatically. The system automatically accommodates accented characters or words in which accented characters are omitted

- Automatic language detection. Real-time translation, however, is not yet available

- Automatic recognition of dates

- Identification of proper nouns, compound words, acronyms, symbols and abbreviations

- Disambiguation of words and phrases. A related operation identifies syntactic information that differentiates homographs such as "leaves" (noun) and "leaves" (verb)

Like many of the university-inspired search engines from the 1990s, Delphes made a number of remarkable claims. For example, a Delphes data sheet called "Delphes Intelligent Knowledge Online Self-Service" asserted that the system could "distinguish particular expressions (locutions) such as "bank on and "bank left." Omitted was the need to maintain dictionaries. Automatic recognition of words and phrases used with fresh connotations was outside the capabilities of Delphes and its competitors. Machine processing of content remains a work in progress despite marketers' assurances that systems can "understand" textual content in a cost-effective, accurate way.

Text: "It keeps desktop, corporate and web content organized so that users do not have to waste time finding, organizing and retrieving valuable information."

Constituent phrase-tree structure:

It keep desktop, corporate and web content organize so that user do not have to waste time find, organize and retrieve value inform

Legend: ▪ head • phrase: S (sentence), NP (noun phrase), VP (verb phrase), SC (small clause), AP (adjective phrase)
• stem • disambiguated tag: N (noun), V (verb), A (adjective), CON (conjunction), NEG (negation, C (complementizer), I (infinitive), AUX (auxiliary)

The Delphes parsing method is similar to the approach taken by other academic systems; for example, the system developed by Syracuse University and later commercialized by Manning & Napier, a financial services firm. The approach is computationally expensive and can be impractical when indexes have to be refreshed with updates for new and changed content.

Role-Based Controls

Compatible with existing security systems, including Basic, NTLM, DPA, Cookie/Script, and HTML/Form authentication methods. User access to content can be controlled via the administrative interface. The profile func-

tion makes it possible to categorize content for a particular user. Permission-based security is supported.

Group, category, and file management levels are supported. Delphes states that its system supports industry standards; nevertheless, Delphes is a proprietary system. Delphes will operate with Dot Net and Internet frameworks.

Knowledge Worker Functions

Delphes stated that it had developed an information manager designed for the knowledge worker. The professional with access to the Delphes system had access to a "dashboard." The dashboard included content creation functions. Delphes explained the dashboard as a way "for creating new, relevant information dynamically." Via the component, a manager could add comments submitted by other system users. A user could boost or lower a result in a hit list. The manager module could be used to "save or share knowledge" by e-mail or by exporting onto CD-ROMs, network folders, or to external applications. The bottleneck in this approach is that one or more individuals have to intermediate functions for the search system.

Intelligent Agent

One of Delphes' most interesting assertions concerned its "Intelligent Agent" for customer service applications. The Delphes system, it is claimed, can "establish a relation between concepts and preferential documents determining its display order." The idea is that content for answering customer service questions can be automatically integrated with relevant content processed by the Delphes system.

This module lets companies personalize their customer service and optimize interaction with clients. Administrators can pre-select one or more documents containing promotional or other targeted information. Consequently, whenever a query relates to a concept matching information within pre-selected documents, those documents will be automatically presented to users in a specific, fully customizable section.

Also, Delphes included a component to make it easy to create a direct mail campaign.

Selected Features

Database

Delphes makes use of a propriety database called eDB. The data management system was designed to improve query processing and results generation. Delphes claimed that its database system was four times faster than other database systems. However, Delphes did not identify the systems against which Delphes was benchmarked. According to Delphes, its data-

base permitted sorting of search results by relevance, date, url, file type, and "section weight and indexing depth." There are desirable features and competitive systems struggled to deliver fast, accurate sorting.

Customization

Customization. The search results page can be customized according to requirements, for example to remain consistent with the look of standard company Web pages. The Delphes interface uses HTML, JavaScript, and Perl scripts. The system designer can:

- Adapt the results page to company colors
- Modify the query box
- Define the number of documents to display on screen
- Change the colors of the contextual document extracts found
- Modify the advanced search criteria
- Change the way you use the search page and select items to display
- The search page, which can be customized.

Fine-Grained Controls

The authorized user compiles data assets in real time and maintains a history of searches and summaries generated by individual users. Features include:

- Dynamic management of informational assets
- History of searches and generated summaries
- Automated classification of each summary based on the search carried out
- Set up profiles and match documents to those profiles

The administrative controls allows the authorized user to set up multi-index queries; that is, passing queries across multiple indexes.

Enterprise Search SDK

Delphes provides a software development kit to licensees. The SDK includes:

- A Web services component for easily integrating and customizing search and summary functionality
- Dot NET, COM and Perl API for easily integrating search or administration tasks into third-party application in order to deliver a search enabled application
- Plug-in API support: This component allows a developer to create a customer search to access content in a data repository. The component can

be used to create connectors and converters so the information can be processed by the Delphes indexing module.

- Support of various programming languages; for example, C++, C, Perl, Visual Basic, C#, Visual Basic Dot NET, Active Server Pages and Active Server Pages Dot Net.

The system also included sample code so the licensee can tap information in Microsoft Exchange Server, a third party content management system, and IBM Lotus Notes.

ArnoldIT Opinion

The Delphes system was extremely ambitious. However, the complexity of the system is a concern for organizations looking for a way to search digital content without making a massive investment in staff, infrastructure, and development.

Table 2: Delphes DioSMW Checklist

Attribute	Delphes Asserts	ArnoldIT Comment
1 Platform	Microsoft Windows	
2 Key word search	Yes	Concept searching is supported
3 Text mining	Yes	Focus is on system usage
4 Automated indexing	Yes	Subject matter experts have to ride herd on user-generated comments and dictionaries
5 Personalization	A developer can customize the system via the Delphes software development kit	The time required to implement varies by developer and organizational requirements
6 Workflow	A developer can customize the system via the Delphes software development kit	The time required to implement varies by developer and organizational requirements
7 Interface	Provides key word search and hot links	Customization permitted
8 Hosted service	Yes, the Dio Web Series is a hosted version of the system	
9 Administrative interface and tools	Graphic and command line interfaces provided	The system can be extended via C++, C, Perl, VB, C#, Visual Basic, Dot Net, Active Server Pages, and ASP.NET.
10 Application programming interface	Yes	Requires the software development kit
11 Professional services	Yes	
12 Security	Supports existing Web security systems	Group and category security supported. File level security supports

Attribute	Delphes Asserts	ArnoldIT Comment
13 Connectors	Connectors are provided for 256 file types	The software development kit makes it possible to access Microsoft Exchange and Lotus Notes content
14 Support for structured data	No	The system can index text records exported from a database. Direct interaction with an enterprise system using structured data must be developed by the licensee
15 Relevance ranking	Yes	Hit boosting and user comments are supported
16 Video	No	
17 Federated search	Yes	
18 Fielded search	No	Sorting is permitted by date and other attributes
19 Content crawler	Yes	
20 Price	Pricing is based on the number of documents the system manages	A collection of one million documents will cost about US$2.2 million

Anticipated Benefits

Delphes makes a number of quite forward-leaning assertions about the benefits of its system. These range from access to "all" information to enabling access to the knowledge an organization has.

Delphes presents its system in terms of lofty concepts. Let me highlight three from Delphes' own documentation:

• Adds value

• Share your knowledge

• Enrich your wisdom

For an organization interested in search and retrieval that combines a wide range of solutions to often ill-defined issues. The idea that knowing who in an organization possesses specific information is important in large organizations. For an organization with a management fully committed to the Delphes approach to information, Delphes may deliver more robust information access than key word retrieval.

Possible Drawbacks

The concerns I have regarding Delphes pivot on the academic nature of the system. Delphes does not deliver search. Delphes delivers a linguistic system that taps content in many different systems and file types. The scope of the Delphes system reaches to content residing on the Internet. Significant

financial, technical, and human resources are required to set up, tune, and maintain the Delphes system.

Other drawbacks include:

- System performance particularly in terms of content acquistion, processing, and indexing

- The need for subject matter experts and system administrators to handle profiles, dictionaries, and user-provided comments (assuming system users take the time to provide comments on documents in the system)

- The complexity of the linguistic model requires a robust infrastructure to handle indexing and concept identification. The UNAX method and parsing model may exceed the capacities of all but the most robust infrastruce.

Delphes is more of an academic solution to the problem of finding knowledge in an organization.

Conclusions

Delphes is similar in its sweeping vision to other search vendors eager to demonstrate that their information retrieval system does more than "search." The problem is that while the vision presented is compelling, the reality does not meet expectations. Many companies that lack deep experience with content processing assume that the marketers' descriptions match the reality of the software.

Delphes--like Convera, Endeca, Fast Search & Transfer, and Verity--gather many different information-centric operations into one "search basket." Neither Delphes nor any of the competitors following a similar path can make a strong case that the value of the deployed system is greater than the total cost of the system.

Delphes has a low profile outside of Francophone nations. The firm's academic and philosophical approach has not found significant traction in the US market. It is unlikely that Delphes will be able to make sufficient market headway to remain in business.

Stephen E Arnold
Minor edits to a rough draft on October 21, 2013